

クラウドソーシングの言語表現：ビジネス日本語研究におけるAI技術の活用

著者	浅井 達哉, 横野 光, 柳瀬 隆史, 岩崎 拓也, 井上 雄太, 田中 啓行, 石黒 圭
雑誌名	専門日本語教育研究
号	22
ページ	9-16
発行年	2020
URL	http://id.nii.ac.jp/1328/00003509/

寄稿

クラウドソーシングの言語表現 —ビジネス日本語研究における AI 技術の活用—

浅井 達哉¹, 横野 光¹, 柳瀬 隆史¹, 岩崎 拓也², 井上 雄太³, 田中 啓行⁴, 石黒 圭²

本論文では、国立国語研究所と富士通研究所が現在行っている、今まで取り扱うことが難しかった、現実のビジネス文書を取り扱った研究を紹介する。そのうえで、これまでの取り組みを発展させた、より緻密な分析を行い、その結果を検証することで、新たなビジネス日本語研究とその貢献の可能性を検討する。具体的には、まず、クラウドソーシングの発注文書を収集し、最新の AI 技術である Wide LearningTMを用いて分析を行った。その結果から得られた興味深い事例をいくつか取りあげ、直観や内省で発見することが難しい、新たな知識が発見できる手続きを提示した。そして、本研究のような分析手法が、今後のビジネス日本語研究にたいして新たな知見を提供できる可能性を示唆した。

キーワード：ビジネス日本語、ビジネス文書、クラウドソーシング、Wide LearningTM、知識発見

1. はじめに

かつて、ビジネス文書研究には大きなハードルが存在した。研究には実際のビジネス現場で使われているビジネス文書が必要である。しかし、社内文書の場合、秘密保持の義務があるために収集することは現実的に不可能であった。そのため、従来のビジネス文書研究で用いられる文書は企業の年次報告書や web サイトなどの社外向けの文書に限られてきた^{1,2)}。多くの研究は、市販のビジネス文例集を用いたり³⁾、調査として状況を設定してビジネスマンにメールを書いてももらったりすることによって行われており⁴⁾、ビジネスの担当者間で交わされる生のテキストを分析すること、とくに量的な側面から分析することは困難であった。

ところが、クラウドソーシングの台頭により、無数の発注文書が、クラウドソーシング運営会社のサイト上で閲覧・収集することが可能となった。クラウドソーシングのビジネス文書は、顔が見えない相手に業務を依頼する発注文書がその中心であり、適切な情報を過不足なく理解できるように書かなければならない。

その発注文書を大量に収集することで、これまで困難であった量的な視点からの分析が可能な状況が整った。

筆者らがこれまで取り組んできた研究と成果は論文集にまとめられている⁵⁾。本論文では、現在試行錯誤を重ねている新たな分析手法を提案するが、まずは、その分析にさいして前提となった、これまでの研究成果の概要を紹介する。

分析観点は、計算機を用いた量的な分析と言語学的な観点による質的な分析に大別される。計算機を用いたビジネス文書の量的な分析として、筆者らは株式会社クラウドワークス^{注1)}のサイト^{注2)}に掲載された発注文書を対象として、最新のデータマイニングや AI の技術を用いた分析に取り組んできた⁶⁻⁸⁾。ビジネス文書の分析においては、分析対象の文書がビジネス上の目的を達した場合に「よい文書」とみなし、目的を達することができなかった場合は「悪い文書」とみなすことができる。こう考えた場合の分析の目的は、「よい文書」と「悪い文書」の違いを明らかにし、「悪い文書」を「よい文書」に改善する方法を示すこととなる。

クラウドワークスの発注文書は、発注する業務の基本情報（業務カテゴリ、発注金額、応募期限、納期など）を示す非言語項目と、発注業務の詳細を日本語で記述した言語表現で構成される。発注文書のビジネス

¹株式会社富士通研究所

²国立国語研究所

³一橋大学大学院言語社会研究科博士後期課程

⁴中央学院大学

上の目的は、発注する業務を請け負ってくれる受注者を獲得することだと考えられる。そこでまず、主に非言語項目を対象として、応募がよく集まる発注文書とそうでない発注文書を弁別する最適な規則群を抽出した⁶⁾。ここでは「個人向けの業務で、予算額が10万円以上であり、逐次支払いを認めないような発注では、発注書のタイトルを短くすると応募が集まりやすい」などの最適な規則を見つけることができた。

次に、言語項目を考慮した分析に着手した。非言語項目だけではなく、言語項目で用いられる品詞の出現数などの簡単な言語属性を加え、富士通研究所の最新のAI技術であるWide LearningTM^{注3}を用いて分析した。その結果、応募率の高い発注文書と低い発注文書を弁別する属性の組み合わせを網羅的に発見し、たとえば「低報酬で応募期間が短い案件では、副詞が多いと応募されにくい」「応募期間が長い場合、仕事内容を簡潔に記載した方が応募されやすい」などの知見を得ることに成功した。

また、言語学的な観点から行われている研究を取りあげると、カッコを中心とした記号の分析⁹⁾、日本語教育の観点から見た文法¹⁰⁾、文章の印象に影響を及ぼす副詞と接続詞^{11,12)}、発注文書において意外と使われるオノマトペ¹³⁾、敬語を中心としたポライトネス¹⁴⁾、発注者の人柄を反映する発話キャラクタ¹⁵⁾と、ビジネス文書の表現に関わる特徴を、さまざまな角度から具体的かつ実証的に論じる試みを行ってきた。

本論文では、これまでの取り組みを発展させ、発注文書において注目すべき重要な言語表現を言語学の観点から選定する。そして、言語表現同士の組み合わせや言語表現と非言語項目の組み合わせを網羅的に分析することにより、「よい文書」と「悪い文書」を弁別する特徴的な組み合わせの発見を目指すものである。

2. 本論文の目的

本論文では、クラウドソーシングの発注文書に着目する。その発注文書の目的は、不特定多数の潜在的な働き手に仕事内容を説明し、多くの応募を求めることにある。つまり、「よい発注文書」には多くの応募が見込まれる一方、「悪い発注文書」には応募数は少なくなるはずである。本論文では、この観点を基準に、

応募数の多寡と、発注文書内で使用される副詞と文末表現といった言語表現との関係性を明らかにする。

具体的には、発注文書のデータベースをクリーニングして類似データの除去を行い、分析にあたって適切なデータとなるように整形する。次に、副詞と文末表現にかんするできるだけ網羅的なリストを作成し、「よい発注文書」との相関を明らかにする。そのうえで、単独の言語表現が発注の成否とどのように関わるのか、また、言語表現を組み合わせた場合と非言語的な情報とがどのように関わるのかそれぞれ分析・考察する。

3. 本論文で使ったデータと分析手法

3. 1 本論文で使ったデータについて

分析に使用した発注文書は株式会社クラウドワークス^{注1}から提供を受けた。このデータはクラウドワークスのサイト^{注2}に投稿された仕事情報のうち2017年8月に投稿されたものである。

データのエントリ数は28,768件であり、各エントリには、仕事のタイトルや内容と言った言語情報と、仕事の作業単価や募集人数、応募期間などのような非言語情報が含まれている。本論文ではこのうち言語情報として仕事の詳細情報、非言語情報として応募期間、報酬、募集人数、仕事カテゴリを利用している^{注4}。

クラウドソーシングにおいて発注者は、募集を行っている仕事にどれだけ多くの作業者が興味を持ち、受注を希望するかを重要視する。この点から本論文では各エントリにたいする応募率を以下の式で定義した。

$$\text{応募率} = \frac{\text{応募者数}}{\text{閲覧回数}}$$

本論文では、この定義に基づき、応募率が高い発注文書を「よい発注文書」、応募率が低い発注文書を「悪い発注文書」とみなす。以降では、データ分析やAI分野における標準的な用語にならって、「よい発注文書」を「正例(Positive example)」と表記し、「悪い発注文書」を「負例(Negative example)」と表記する。

3. 2 データのクリーニング

分析データは28,000件超の大規模なものであるが、実際の仕事内容には、ある発注者が似たような商品にかんするレビュー執筆を商品毎に別々の発注として投稿しているものがある。この場合、仕事内容が商品部

分以外は共通していることが多く、結果として非常に似たテキストが多数存在することになる。これが頻度に基づいた分析に悪影響を及ぼすおそれがある。この影響を低減するために、文面が似たエントリを削除することでデータのクリーニングを行った。

一般には、各仕事内容間の類似度に基づいたクラスタリングなどを行い、各クラスタから事例をサンプリングすることでクリーニングを行うことが考えられる。しかし、本データでは内容が似ていても応募率が異なるという状況が発生する。当該の仕事が投稿された時期によって比較される他の仕事が異なっていたり、商品レビューの場合、商品の知名度・普及度によって受注可能かどうか異なっていたりするからである。

そこで、応募率と内容の類似度を用いてデータのクリーニングを行った。具体的な処理は以下に示す。この処理によって最終的に得られた正例 2,802 件と負例 3,250 件を分析対象とする。

1. 各エントリの応募率を計算し、上位 25%と下位 25%のエントリの応募率をそれぞれ正例の閾値、負例の閾値とする
2. 各エントリ間の類似度を計算する。類似度は仕事内容の単語 bi-gram によるコサイン類似度を用いる
3. 類似度が 0.95 以上のエントリ同士をまとめることでクラスタリングを行う
4. 各クラスタ内のエントリの応募率の平均を求め、その平均が正例の閾値以上か負例の閾値以下であれば、そのクラスタから代表とするエントリを 1 件抽出する。応募率の平均値が条件を満たさないクラスタは分析データから削除する

3. 3 Wide Learning™による「知識発見」

3. 3. 1 Wide Learning™について

Wide Learning™^{注3}は、学習データの中からデータ項目の組み合わせで表現された重要な仮説（ナレッジチャンク）を高速に列挙し、その知識を用いて予測・分類を行う技術である^{7,8)}。Wide Learning™は、人間が理解可能な仮説を用いているため、出力結果の根拠を説明できる「説明可能な AI(Explainable AI; XAI)」である

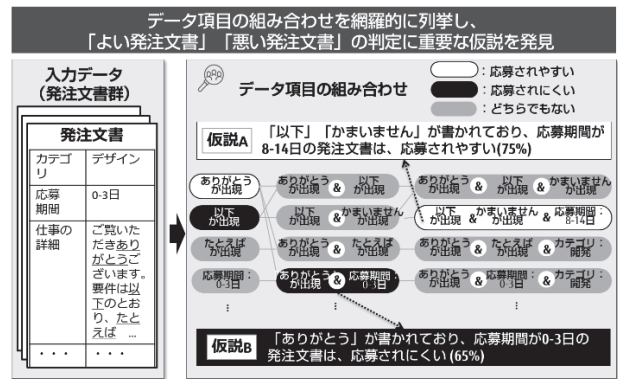


図1 Wide Learning™によるビジネス文書分析

点が特長の一つである。また、与えられた制約条件を満たす仮説空間を網羅的に探索し、重要な仮説を漏れなく列挙するので「知識発見」が可能である。上の図1に、本研究の Wide Learning™の適用イメージを示す。

計算機を用いた一般的な言語分析では単語の出現頻度や単語の共起の相関などから特徴的な要素を発見するが、非言語項目やより多くの要素の組み合わせを考慮すると組み合わせの数が非常に多くなり分析が困難となる。この問題にたいして機械学習モデルで学習を行い、素性に対する重みなどによって分析を行う言語処理的なアプローチがあるが、得られた結果に対する説明性が問題となることがある。これにたいして、我々の研究では「知識発見」を重視する立場からのアプローチをとる。

3. 3. 2 Wide Learning™を用いた分析

前述の通り、分析データには言語情報と非言語情報

表1 分析に使用した言語表現(一部)

大分類	小分類	事例(一部)
発注者側	一人称	わたし、私、当方
	発注者の行為	いたします
受注者側	二人称	あなた、そちら
	受注者の行為	ご応募ください
社交表現	感謝	ありがとう、感謝です
	謝罪	失礼しました
接続表現・参照表現	接続表現	そのため、しかし
	参照表現	以下、上記
程度量副詞	量副詞	おおよそ、すこし
	頻度副詞	たまに、よく

表2 非言語情報の分析項目

種類	項目
募集人数	1人、2-5人、6-10人、11人以上
応募期間	0-3日、4-7日、8-14日、15日以上
報酬 (下限)	5,000円未満、5,000円、10,000円、50,000円、 100,000円、300,000円、500,000円
カテゴリ	ライティング、開発、デザイン、ビジネス事務

が含まれている。言語情報は、あらかじめ文書の印象に影響を与えると考えられる言語表現を283件ピックアップし、それらが仕事内容に出現しているかをもとに分析を行った。その一部を前ページの表1に示した。

一方、非言語情報は、応募期間、報酬、募集人数、カテゴリを分析対象とした。応募期間など数値で表される表現はそれぞれ設定した範囲で扱う。本論文では上掲の表2のとおり19種類の項目を用いる。このデータにたいしてWide Learning™を適用する。

本論文では、以下の二つの制約条件で規定される仮説空間を網羅的に探索して得られたナレッジチャンクを考察対象とする。

- ・正例と負例の合計出現数：5回以上
- ・組み合わせの最大数：3

一つ目の制約条件は、ほとんど発生しない例外的な事例からは、一般に活用可能な有益な知識が得られないので、本論文での検討の対象外とすることを表す。二つ目の制約条件は、本論文では小さい組み合わせから検討を開始し、全体的な傾向を把握することを意図している。将来、本論文での検討結果を発展させる過程で、ここで与えた組み合わせの最大数を増やして再分析を行うことで、さらに新しい知識の発見も可能となる。

以上の制約条件のもとでWide Learning™を動かし、得られたナレッジチャンクの総数は1,319件であった。4節では、これらのナレッジチャンクを、「単独の言語表現」「2種類の言語表現の組み合わせ」「言語表現と非言語項目の組み合わせ」に分け、それぞれ重要なものや興味深いものをいくつか抽出して考

察を行う。

4. 分析結果

4. 1 言語表現とラベルの相関

組み合わせを分析する前に、まずは単一の言語表現と発注文書の良し悪しとの関係性を分析する。Wide Learning™では、長さ1のナレッジチャンクとしてこれらの分析が可能である。

3. 1で定義した正例（「よい発注文書」）と負例（「悪い発注文書」）において、本論文の分析に使用した言語表現がどちらで多く使用されているのかを知るために、正例数と負例数、相関係数を求めた。表3に相関係数の上位15例を示す。ここには、正の相関が強い（すなわち、「よい発注文書」に出現しやすい）言語表現が順に並んでいる。同様に、表4に相関係数の下位15例を示す。ここには、負の相関が強い（すなわち、悪い発注文書に出現しやすい）言語表現が順に並んでいる。

なお、通常の実データ分析では、相関係数が0に近い（すなわち、正の相関も負の相関もほとんど認められない）データ項目を除外することがよく行われる。しかし、単一の項目としては相関が認められなくても、他の項目と組み合わせることで正または負の相関が顕在化することがある。そのような組み合わせを見逃さずに発見するために、本論文では3. 3. 2で示したすべての言語表現と非言語項目を分析に用いた。

以降では、次ページの表3と表4のそれぞれの特徴をもとに考察を行う。まず、文末表現に着目してみると、正例では、「お待ちしております」「致します（いたします）」「ご応募ください」といった表現が使用されていた。一方、負例では、「してください（して下さい）」「禁止です」「お願いします」「できません」といった表現が使用されていた。これらの表現を比べてみると、正例では「～ております」「致します」のような発注者が受注者にたいして一歩引く、謙譲的な配慮表現が好まれることがわかる。これにたいし、負例では「お願いします」「してください」のような丁寧体ではあるが、相手の意向を確かめずに一方的に依頼する、配慮に欠けた言語表現が多いと言える。

表3 「良い発注文書」に見られる特徴的な言語表現

順位	言語表現	正例数	負例数	相関係数
1	お待ちしております	1081	313	0.343
2	気軽に	828	322	0.250
3	たくさん	667	231	0.234
4	こちら	855	406	0.221
5	コツコツ	122	19	0.125
6	弊社	510	315	0.124
7	以降	182	63	0.115
8	致します	670	487	0.113
9	いただければ	384	226	0.112
10	いたします	1037	909	0.097
11	しっかり	165	71	0.095
12	尚	105	30	0.095
13	ご応募ください	149	60	0.095
14	丁寧に	88	22	0.092
15	主に	112	41	0.087

表4 「悪い発注文書」に見られる特徴的な言語表現

順位	言語表現	正例数	負例数	相関係数
186	あなた	126	419	-0.146
185	してください	669	1193	-0.139
184	必ず	180	481	-0.134
183	禁止です	63	220	-0.107
182	して下さい	121	306	-0.099
181	以下	479	789	-0.088
180	自由に	48	150	-0.081
179	以上	633	964	-0.080
178	なぜ	18	68	-0.061
177	お願いします	861	1185	-0.060
176	どう	104	202	-0.057
175	できません	110	206	-0.054
174	かならず	2	22	-0.048
173	?	232	349	-0.042
172	めっちゃ	0	12	-0.041

発注文書の性質は、発注者が受注者より圧倒的に強く、不均衡な力関係のもとに存在するものであるとさ

れる¹⁶⁾。そのため、この不均衡な力関係をやわらげるために、発注者は自らのキャラクタを「よい発注者」にしなければならない¹⁵⁾。そこで、言語表現は、言い回しを丁寧にするだけでなく、配慮を込めることで「よい発注者」「よい発注文書」につながると考えられる。

つぎに、副詞に着目してみると、正例では、「たくさん」「コツコツ」「しっかり」「丁寧に」といった、発注者が求める受注者像にかんする表現が使用されていることがわかる。ところが負例では、受注者にたいして全く指示がない「自由に」といった表現や、反対に「必ず（かならず）」のような強い限定を伴う表現が使用されていた（文末表現の「禁止です」や「できません」も強い限定を伴う表現である）。このような副詞の使用は、使い方次第で相手に不快感や嫌悪感を与えることにつながり、負例の言語表現に特徴的な一例であると考えられる。なお、「主に」のように受注者への指示の大枠を示す副詞や、「気軽に」のように業務の限定性を弱め、受注者の新規参入を促す副詞が、正例に特徴的な言語表現として現れるのも、この裏付けであると考えられる。

しかしながら、単純になぜその言語表現が正例（または負例）の特徴として現れたのか説明ができないものも存在する。参照表現を例にとると、「以降」が「正例で」「以上」「以下」が負例の特徴として現れているが、これがなぜなのか、この段階では説明がつかない。そこで、次節では、複数の言語表現を組み合わせた結果と言語表現と非言語項目を組み合わせてみた場合の結果をもとに考察を行う。

4. 2 組み合わせに着目した場合

4. 2. 1 言語表現の組み合わせ

ここでは、Wide LearningTMを適用し、発見したナレッジチャンクをもとに得られた言語表現の組み合わせについて考察していく。まず、正例において特徴的であった言語表現が、言語表現の組み合わせによっては負例における特徴的な言語表現になってしまった例を取りあげる。

表3において「致します（いたします）」は正例における特徴的な言語表現であった。しかし、同様に正例において特徴的な言語表現である「是非」を組み合

わせると、負例の特徴的な言語表現になることが明らかになった。以下の(1)(2)に実際の例を挙げる。

このような例は、「是非」に限らず、正例に多く出現していたほかの副詞「何卒」や「どうぞ」においても確認された((3)(4)を参照)。これらの副詞はともに「よい発注文書」に特徴的な言語表現である。

- (1) 是非、よろしくお願いいたします。
- (2) 是非協力よろしくお願いいたします。
- (3) 何卒、宜しく願い致します。
- (4) どうぞ、宜しく願い致します。

このような強調する副詞をお願いする表現とともに使用すると、業務量の多い仕事や難度の高い仕事を押しつけるような印象を与え、受注者に好まれなくなると考えられる。

次に、負例において特徴的であった言語表現が、言語表現の組み合わせによって正例における特徴的な言語表現になった例を取りあげる。「できません」は4.

1でも確認したように、単体では負例における特徴的な言語表現であった。しかし、同様に負例に多く出現していた「予め(あらかじめ)」と組み合わせることによって正例に特徴的な言語表現となる。実例である(5)と(6)を見てみると、「できませんので予めご了承ください。」のように、できないことを単に伝えるだけでなく、前もって知っておいて、理解してほしいという気持ちを表出している。このように、ナレッジチャンクから導き出された組み合わせから母語話者でも気づけなかった新たな表現と文書の良し悪しとの相関を発見することができる。

- (5) 逆に不採用の場合は、テストライティング分のお支払いはできませんので予めご了承ください。
- (6) ※応募サイトのワーカー名の回答を間違えますと、本人様の確認ができないためご案内でできませんので予めご了承くださいますようお願い致します。

4. 2. 2 言語表現と非言語項目の組み合わせ

最後に、言語表現と非言語項目の組み合わせとして

表5 言語表現と非言語項目の組み合わせ例(1)

単一項目と組み合わせ	正例数	負例数
[言語表現]ありがとう	492	374
[非言語項目]応募期間 0-3 日	399	417
[言語表現]ありがとう & [非言語項目]応募期間 0-3 日	43	77

ナレッジチャンクについて、興味深いものをいくつか紹介する。

まずは、上の表5に示したナレッジチャンク「[言語表現]ありがとう & [非言語項目]応募期間 0-3 日」を説明する。言語表現の「ありがとう」は、仕事の発注者から、発注文書を閲覧してくれた人への感謝の気持ちを表しており、一般的には丁寧で望ましい表現だと考えられる。実際に、「ありがとう」と書かれた発注文書は、正例の出現数の方が負例への出現数より3割程度多くなっている。一方、非言語項目の「応募期間 0-3 日」は、応募期間が短いことを表している。正例と負例の差はそれほどない。

ところが、これらを組み合わせて、「ありがとう」と「応募期間 0-3 日」が両立すると、負例の出現数が正例の出現数より8割程度も多い結果となった。すなわち、応募期間が短い仕事の発注文書においては、「ありがとう」という感謝の表現を書くと、応募が減少することがわかる。

具体的な発注文書を見ながら理由を考察してみよう。クラウドソーシングのウェブサイトでは仕事を探す人は、検索を繰り返しながら数多くの発注文書を見比べて、自分のやりたい仕事に応募する。応募期間が短い仕事の発注文書にアクセスする人は、次の仕事を急いで確定させたいので要件のみを速やかに把握したいことが想定される。感謝の表現が書かれた丁寧な発注文書は、その後に書かれているであろう要件についても言葉を尽くして丁寧に書かれており、内容の把握に時間がかかり、かつ業務負担量も相当な量であることが予想される。その結果、仕事を探している人が、その文書を読み終える前に次の発注文書に移ってしまうことが考えられる。感謝の表現から開始される丁寧な発注文書(応募期間 0-3 日)の事例を次ページの(7)に示す。

表6 言語表現と非言語項目の組み合わせ例(2)

単一項目と組み合わせ	正例数	負例数
[言語表現]以下	479	789
[非言語項目]ビジネス事務	677	929
[言語表現]以下 & [非言語項目]ビジネス事務	124	139

(7) はじめまして、【発注者名】と申します。

数ある募集からご覧いただきありがとうございます。

募集内容については下記をご確認いただきますようお願い致します。

次に、上の表6に示したナレッジチャンク「[言語表現]以下 & [非言語項目]ビジネス事務」を説明する。

「以下」は、数量や文章の参照を表す言語表現であり、文章の参照を表す場合は、フォーマルで硬い印象を与える表現でもある。今回のデータでは、「以下」と書かれた発注文書は、負例の方が正例よりも6割以上多く出現していることから、応募されにくい言語表現の一つとなっている。「以下」という言語表現に続けて、要件を無味乾燥に並べる書き方が、全体的傾向として応募されにくい一因となっている可能性がある。

また、非言語項目の「ビジネス事務」は、募集する仕事のカテゴリの一つであるが、今回のデータでは負例への出現数が正例よりも3割以上多く、応募されにくい案件となっている。

ところが、この二つが組み合わせり、「ビジネス事務」に「以下」が出現すると、正例と負例の差が大幅に縮小した。応募されにくいはずの二つの項目が組み合わせれば、一層応募されにくくもおかしくないはずなのに、これはどういうことだろうか。

その理由として、「ビジネス事務」という仕事カテゴリの発注文書と「以下」のようなフォーマルで硬い言語表現との相性が悪くないことが考えられる。ビジネス事務の応募要件を伝える事例を(8)に示す。

(8) 基本的には指示通りの手順で進めていただければ問題ありませんが、以下の知識があると作業がスムーズです。

- ・スプレッドシート（エクセル）の関数の知識（split 関数や unique 関数を使用します。）
- ・簡単なテキストエディタの操作方法（ショートカットコマンドなど）の知識

このように、Wide Learning™ではあらゆる組み合わせの網羅的な列挙により、興味深い仮説や意外な仮説を見つけることができる。なお、筆者らは3項目からなるナレッジチャンクも検討し、いくつかの発見も得ているが、紙幅の都合上、別の機会に紹介したい。

5. まとめ

本論文では、従来取り扱うことが困難であった現実のビジネス文書を対象に進めてきた筆者らのこれまでの取り組みを発展させ、言語表現や非言語項目の組み合わせを網羅的に分析することで、「よい発注文書」と「悪い発注文書」を弁別する特徴的な組み合わせの発見を目指した。

こうした量的なデータを扱うさいには、重複した文面を削除することが重要になる。そこで、本論文では、データのクリーニングを最初に実行したうえで、最新のAI技術であるWide Learning™による分析を行った。これにより、直観や内省では発見することが難しい新しい知識の発見を試みた。そのうえで、本論文では、用いたデータ項目の組み合わせから示唆されたナレッジチャンクを「単独の言語表現」「2種類の言語表現の組み合わせ」「言語表現と非言語項目の組み合わせ」に分け、それぞれ重要だと思われるものや興味深いものを取りあげて考察を行った。

まず、「単独の言語表現」では、丁寧にするだけでなく、配慮が示されていることが大切であり、それを表すような言語表現を使用しなければ「よい発注文書」にはならないことが示唆された。

次に、「2種類の言語表現の組み合わせ」では、「よい発注文書」に特徴的な言語表現同士でも、組み合わせによって「悪い発注文書」の特徴になってしまう例と「悪い発注文書」に特徴的な言語表現同士が組み合わせられて「よい発注文書」の特徴になりうる例を取りあげた。これらは、日本語母語話者の直観からでは導きづらい例であり、日本語の研究においてナレッジ

ヤンクを活かすことができる一例であると言える。

さらに、「言語表現と非言語項目の組み合わせ」では、応募期間や文書カテゴリといった非言語項目と言語表現の組み合わせにより、正例が負例に（または負例が正例に）なるという興味深い結果を取りあげた。

このように、生のビジネス文書を最新の AI 技術を用いて分析することは、今後のビジネス日本語研究にたいして新たな知見を提供できる可能性を秘めている。

今後の課題について、いくつか挙げておく。まず、データ整備を進め、より緻密な分析を行うことである。そのなかには、言語表現の出現間隔や出現パターンの考慮、および他の非言語項目も含めたより複雑な関係性の分析などがある。また、言語学的な観点においては、漢字表記にするかひらがな表記にするかといった日本語の表記や、箇条書きや文長、段落構成などの文章構成、といった、従来では量的な分析が困難であった観点を取り入れることを考えている。

謝辞

本論文は、国立国語研究所機関拠点型基幹研究プロジェクト「日本語学習者のコミュニケーションの多角的解明」（プロジェクトリーダー：石黒圭）の研究成果であり、また、JSPS 科研費 17K18504 の助成を受けたものです。また、株式会社クラウドワークスには、無償にて貴重な資料の提供をいただきました。心から感謝申し上げます。

注

注 1 株式会社クラウドワークスは、ランサーズ株式会社と並ぶ、国内最大級のクラウドソーシング運営会社。当社の 2020 年 9 月期 通期決算説明資料によれば、67.0 万社のクライアントと 410.4 万人のクラウドワーカーの登録があるとされる。

注 2 <https://crowdworks.jp/>

注 3 Wide Learning™ 公式サイト「Hello, Wide Learning!」
<https://widelearning.labs.fujitsu.com/>

注 4 具体的な事例は、サイトに掲載されているものを参照されたい。

参考文献

- 1) 佐野彩子：企業の年次報告書を用いたビジネス分野の外來語に関する一考察：アカデミックジャパニーズ、白書、新聞語彙との比較を中心に、専門日本語教育研究, 18, pp.37-42 (2016)
- 2) 諸星美智直：ビジネス文書における副詞「あらかじめ」の用法：企業ウェブサイトと内容証明郵便を中心に、国学院大学日本語教育研究, 5, pp.1-14 (2014)

- 3) 加藤恵梨：接続詞「については」の機能と意味について、大手前大学論集, 19, pp.109-125 (2019)
- 4) 黎虹虹：ビジネスメールにおける日本語の対人配慮の示し方：謝罪表現とその意識を中心にして、昭和女子大学大学院 言語教育・コミュニケーション研究, 10, pp.61-76 (2015)
- 5) 石黒圭編：ビジネス文書の応用言語学的研究 クラウドソーシングを用いたビジネス日本語の多角的分析, ひつじ書房(2020)
- 6) 石黒圭・岩田一成・蒙韞・青木優子・浅井達哉：クラウドソーシングを用いたビジネス発注文書の質的分析試論, 2018 年度日本語教育学会秋季大会
- 7) 大堀耕太郎・浅井達哉・岩下洋哲・後藤啓介・重住淳一・高木拓也・中尾悠里・穴井宏和：知識発見によって信頼をつなぐ Wide Learning 技術, 雑誌 FUJITSU 2019 年 9 月号, 70(4), pp.48-54 (2019)
- 8) H. Iwashita, T. Takagi, H. Suzuki, K. Goto, K. Ohori, H. Arimura : Efficient Constrained Pattern Mining Using Dynamic Item Ordering for Explainable Classification, arXiv, CoRR abs/2004.08015 (2020)
- 9) 岩崎拓也：記号の使用実態とその問題点 発注者と受注者をつなぐためのカッコの活用, ビジネス文書の応用言語学的研究 クラウドソーシングを用いたビジネス日本語の多角的分析, ひつじ書房, pp.79-100 (2020)
- 10) 岩田一成：文法項目の使用実態とその問題点 外国人にも読みやすい日本語は？, 石黒圭編⁵⁾, pp.101-119 (2020)
- 11) 佐野彩子：副詞の使用実態とその問題点 仕事内容を正確に伝える副詞の使い方, 石黒圭編⁵⁾, pp.121-136 (2020)
- 12) 佐野彩子：接続詞の使用実態とその問題点 接続詞は使わないほうがいいのか？, 石黒圭編⁵⁾, pp.137-152 (2020)
- 13) 赫楊：オノマトペの使用実態とその問題点 オノマトペはビジネス文書に不要？, 石黒圭編⁵⁾, pp.153-168 (2020)
- 14) 市江愛：条件表現の使用実態とその問題点 どの表現が丁寧な印象を与えるのか？, 石黒圭編⁵⁾, pp.169-182 (2020)
- 15) 井上雄太：発注文書から伝わる書き手の発話キャラクターとその問題点 ビジネス文書に透けて見える書き手像, 石黒圭編⁵⁾, ひつじ書房, pp.201-216 (2020)
- 16) アンドレイ・ベケシュ：クラウドソーシング発注文書の特徴 発注者と受注者の力関係との関連, 石黒圭編⁵⁾, pp.277-294 (2020)